



ELSEVIER

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

The impact of integrating ICT with teaching: Evidence from a randomized controlled trial in rural schools in China



Yu Bai ^{a, b}, Di Mo ^{c, d, *}, Linxiu Zhang ^a, Matthew Boswell ^c, Scott Rozelle ^c

^a Center for Chinese Agricultural Policy, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, No. 11-A Datun Road, Chaoyang District, Beijing 100101, PR China

^b University of Chinese Academy of Sciences, PR China

^c Stanford University, USA

^d University of Leuven, Belgium

ARTICLE INFO

Article history:

Received 1 August 2015

Received in revised form 26 January 2016

Accepted 11 February 2016

Available online 13 February 2016

Keywords:

ICT in education

Computer assisted learning

Computer assisted instruction

English learning

Randomized controlled trial

ABSTRACT

Recent attention has been placed on whether integrating Information Communication Technology (hereafter, ICT) into education can effectively improve learning outcomes. However, the empirical evidence of the impact of programmes that adopt ICT in schooling is mixed. Theory suggests it may be due to differences in whether or not the ICT programmes are integrated into a teaching programme of a class. Unfortunately, few empirical studies compare the relative effectiveness of programmes that integrate ICT into teaching with the ones that do not. In order to understand the most effective way to design new programmes that attempt to utilize ICT to improve English learning, we conducted a clustered randomized controlled trial (RCT) with some schools receiving ICT that was integrated into the teaching programme of the class; with some schools that received ICT without having it integrated into the teaching programme; and with other schools being used as controls. The RCT involved 6304 fifth grade students studying English in 127 rural schools in rural China. Our results indicate that when the programme is integrated into the teaching programme of a class it is effective in improving student test scores relative to the control schools. No programme impact, however, is found when the ICT programme is not integrated into the teaching program. We also find that when ICT programmes are integrated into teaching, the programmes work similarly for students that have either high or low initial (or baseline) levels of English competency. When ICT programmes are not integrated with teaching, they only raise the educational performance of English students who were performing better during the baseline.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Integrating Information and Communication Technology (ICT) into the classroom is a promising approach to help disadvantaged students across the world (Bakar, Inal, & Cagiltay, 2006; Ebner & Holzinger, 2007; Pal, Pawar, Brewer, & Toyama, 2006; Banerjee, Cole, Duflo, & Linden, 2007; Lai, Luo, Zhang, Huang, & Rozelle, 2011; Mo, Zhang, Luo, et al., 2014). Policymakers in developed and developing countries have tested various methods to use ICT for innovation and improvement

* Corresponding author.

E-mail address: dimo@stanford.edu (D. Mo).

of poor performance students (Sweet & Meates, 2004). In many parts of the world, large investments have been made to integrate ICT into education systems in order to support remedial learning (Butcher, 2011; Zhang, Zhang, Zhao, & Wu, 2014). For instance, Turkey spends 11.7% of its education budget on ICT (Hismanoglu, 2012). China's government has established a plan to increase the investment in ICT in education, with special attention given to remote rural areas (MOE, 2012, 2012, 2014). According to this plan, a computer classroom is to be set up in every rural primary school during the 12th Five-Year Plan (2011–2015—MOE, 2012).

Despite the increasing use of ICT for helping disadvantaged students in schools, researchers have found large heterogeneity in the impact of ICT programmes on student academic achievement. Previous studies in both developed and developing countries have shown inconsistent evidence of the educational benefits of using ICT in schools (e.g. Banerjee et al., 2007; Barrow, Markman, & Rouse, 2008; Almekhlafi, 2006; Olibie, 2010; Lai et al., 2012, 2013; Mo, Zhang, Wang, et al., 2014; Angrist & Lavy, 2002; Rouse & Krueger, 2004; Hlas & Vuksanovich, 2007; Dynarshki et al., 2007; Chien, 2011). For example, Banerjee et al. (2007) showed that a Computer Assisted Learning (CAL) programme significantly helped students increase math scores. Additionally, these increases were sustained when students' educational performance was measured one year after the programme concluded. He, Linden, and MacLeod (2008) and Mo, Zhang, Wang, et al. (2014) have also found positive impacts of ICT programmes in developing countries. In contrast, Angrist and Lavy (2002) came to fundamentally different conclusions, finding that when ICT was used inside the classrooms of Israeli elementary schools, it led to negative, but insignificant, impacts on 8th grade students. Similarly, Rouse and Krueger (2004) demonstrated that an ICT programme in US schools had no significant effect on student reading.

While there are many factors that may have contributed to the heterogeneity of the outcomes, one potential reason for the observed differences in ICT programmes may be that some integrated ICT into teaching (henceforth, *computer assisted instruction* or *CAI*) while others did not. In the rest of the paper we call programmes that do not integrate ICT into teaching *computer assisted learning* or *CAL* programmes.

Whether or not ICT should be integrated into the teaching programmes of instructors is an ongoing debate in the literature (Bingimlas, 2009; Granger, Morbey, Lotherington, Owston, & Wideman, 2002; Mumtaz, 2000; Stockwell, 2007; Tondeur, Valcke, & Van Braak, 2008). On the one hand, many scholars believe that CAI is an effective way to utilize ICT in schools. Researchers have argued that ICT's value in education is that it provides the means to make teaching more efficient and interesting (Hughes, 2005). However, it is believed ICT, while potentially effective, cannot replace the role of teachers in providing timely feedback to their students (Dina & Ciornei, 2013; Hismanoglu, 2013). Therefore, it is not only desirable, but also necessary for ICT to be integrated into the teaching process.

In contrast, there are also proponents of using CAL approaches to education in schools. The main benefit of a successful CAL programmes (if they are indeed successful) is that they can replace teachers and still help students to learn. CAL programmes can also be rolled out in an inexpensive manner due to limited human resource requirements. In this way, CAL programmes are not constrained by factors such as the quality of teaching or teacher effort. Rahimi and Yadollahi (2011) suggest that one of the main issues associated with adopting and integrating ICT into teaching curricular is that teachers may act in a way that undermines the effectiveness of ICT. Teachers may simply choose to use ICT materials without matching the materials to the learning levels of students. If teachers do not expend enough effort to incorporate ICT into their teaching, student performance could potentially be harmed.

Another important point made by the literature is that the effectiveness of different ICT programmes may depend in part on the learning level of students at baseline. For example, Huang et al. (2014) show that the effectiveness of ICT programmes can depend on the initial learning level of students. According to Krashen (1982), calibrating the additional amount of new learning material relative to the student's current level of knowledge is a crucial step to better learning. In other words, the provision of new learning material is most effective when it is provided at a level just beyond the student's current level of competence. As a result, ICT-based instruction needs to be readily comprehensible to students for there to be effective learning (Hatch, 1978; Long, 1996). Such a theory implies that students should at least have a minimum level of knowledge of the subject before ICT can benefit them. In other words, ICT is not a technology that can be effectively used by a student to learn a subject that is totally new to him/her.

In addition, the type of teacher implementing an ICT programme may play a role in determining the programmes' effectiveness. Burston (1996), Jones (2001) and Yang and Huang (2008) show that implementing an ICT programme may increase teachers' workload, since they need to make efforts to prepare for and organize the new programme. Therefore, a teacher's workload at baseline could affect the effectiveness of an ICT programme. If a teacher has a heavy workload, he/she may not be willing to carefully or fully implement an ICT programme, while a teacher with a lighter workload would be more inclined to implement the program. However, researchers have also suggested that if an ICT programme is effective in helping student learn and complements regular teaching, it may require less marginal effort from the teachers to improve student performance (Lam, 2000; Nomass, 2013). In other words, the overall teacher effort may decrease if the ICT programme can help improve student performance, thus improving overall efficiency of teaching. As a result, determining which type of ICT programmes works in a particular context may depend on the nature of the teacher and how well the ICT programme complements teaching practices.

If this theory is true, then we may be able to create a testable hypothesis: the effectiveness of ICT programmes depends on both the nature of students and the engagement of teachers. Under this hypothesis, when an individual student is performing poorly we would expect that CAI programmes would be more effective than CAL programs. Because CAI is rolled out as part of a teaching program, an instructor could help poorer performing students catch up with the subject so that they may better

utilize the ICT program's materials. In contrast, because CAL requires less engagement on the part of instructors (beyond the management of the ICT program), poorer performing students might not be able to keep up with the material being presented by the ICT program.

There is a distinct lack of empirical studies that investigate which types of ICT programmes are most effective for certain types of students or teachers. Many studies that look at ICT programmes in education do not distinguish between programmes that more closely resemble CAL or CAI practices in nature (Angrist & Lavy, 2002; Linden, 2008). Among the studies that do explicitly test the impact of CAI, many are small scale. For instance, a study in Nigeria only involved a total of 160 students in 4 classes (Olibie, 2010). In addition, many of these studies do not have a valid control group or an appropriate identification strategy to determine the impact of CAI/CAL programmes (e.g. Rahimi & Yadollahi, 2011; Almekhlafi, 2006; Hughes, 2005). Notably, to our knowledge there has been no large, empirical study that compares the relative effectiveness of CAI and CAL programs—especially in the context of a developing country. Fewer studies have been conducted to understand which types of ICT programmes are more effective for different types of students or teachers.

The overall goal of this study is to understand whether CAI programmes are effective in improving learning outcomes and to compare the relative effectiveness of a CAI programme against a CAL program. We also are interested in which types of ICT programmes (that is, CAI or CAL) are most effective at helping different types of students and which types are more effectively used by different types of teachers. To meet this goal, we have three specific objectives. First, we measure the impact of ICT programs, in general, on student academic performance. Second, we compare the impact of a CAI programme to that of a CAL programme as a way to investigate the relative effectiveness of CAI. Third, we explore the potential mechanisms through which CAI programmes may be more or less effective than CAL programs. We meet this third objective by demonstrating which types of ICT programmes are most effective for various types of students or teachers.

The rest of the paper is organized as follows: the next section presents the study's sampling, data and methods. The third section discusses the results. We draw conclusions in the final section.

2. Sampling, data and methods

In order to achieve our objectives, we implemented a clustered randomized controlled trial (RCT) in northwest China. A total of 6304 fifth grade students in 127 elementary schools in Haidong prefecture participated in our study. The programme lasted for the 2013–2014 academic year. In the following subsections, we present the study's sampling protocol, randomization procedures, intervention, data collection approach and statistical methods.

2.1. Sampling and randomization

We followed several steps to choose our sample. First, to focus our study on students from poor rural areas, we restricted our sampling frame to Haidong Prefecture, a poor minority area in Qinghai Province in northwest China. Among 31 provinces in mainland China, Qinghai ranked 30th in terms of GDP in 2013 (National Bureau of Statistics, 2014). The annual per capita net income of the selected prefecture was only around 6150 RMB (approximately 990 US dollars) (National Bureau of Statistics, 2014).

Second, after choosing Qinghai province, we chose the sample counties. All six counties in the prefecture were selected to be included in our sampling frame. Five of the counties are nationally designated poor counties (National Bureau of Statistics, 2014).

Third, after choosing the counties, we chose the sample schools. The first step in creating our school sample was obtaining a comprehensive list of schools in the six counties from each county's local education bureau. In order to run the CAI or CAL programmes in the schools, we restricted our sample to schools that met the minimum requirement of computer facilities. The minimum requirement was a ratio of computers to class size of at least 0.25. This ratio was needed to ensure that at least two students would be able to share one computer (that is, there were enough computers so each school had enough computers to at least allow a class of students to break up into two groups and participate in the program). We found a total of 130 schools in the six sample counties that meet the requirement.

Within our sample schools, the intervention was focused on fifth graders. According to the national curriculum, primary students are supposed to start learning English in the third grade (MOE, 2001). However, due to a lack of English teachers, many schools in China's minority areas do not start teaching English until the fourth or fifth grade (Zheng, 2009). Because of this, we decided to choose fifth grade for the study. This choice was thought to allow us to be certain that all students would have begun to learn English in school and that all schools (at least for grade 5) had English teachers.

Before selecting our sample schools, we employed power calculation to make sure there were enough sample schools for evaluating the impact of our programme in this research. In the power calculation, we assumed a standardized effect size for the outcome variable of 0.20, 0.80 power, a five percent significance level, an intra-cluster correlation (ICC) of 0.1, and a pre- and post-intervention correlation of 0.5. We also assumed that there were 50 observations in each cluster, on average. Based on these assumptions, we calculated that we needed a total of 128 schools to be included in our experiment. Given a limited budget to implement the project and a cost ratio of 1:2 between each control unit and each treatment unit (that is, the expected costs to be spent in each treatment school was about twice as high as those in each control school), we randomly assigned 22 schools (each) to each of the two intervention arms and 84 schools to the control arm (Duflo, Glennerster, & Kremer, 2007).

Based on our power calculations we randomly chose 128 schools and divided them into three groups: two treatment groups and one control group. We randomly assigned 22 schools to receive a CAI intervention (the exact nature of the treatment is described in detail below). We also randomly assigned another 22 schools to receive a CAL treatment. The rest of the 84 schools were assigned to the control group. We had to drop one school (from the control group) that did not have a teaching staff that met our study's requirement. Ultimately, there were a total of 1236 (19.6%) students in the 22 CAI treatment schools and 1068 (16.9%) students in 22 CAL treatment schools. The rest of 4000 (63.5%) students in 83 schools served as the control group.

Our randomization process successfully created a balanced sample across the treatment and control groups. To show this, we used a set of student and school characteristics to analyze whether there were any significant differences among the three groups. In doing so, we regressed the baseline characteristics on the treatment dummy variables. The results show that none of the variables are significantly different among the three groups (Table 1). Therefore, we can say that there were no imbalances between the treatment and control groups among those variables at baseline. In addition, all the differences between treatment and control groups are small in magnitude.

Although at the time of the baseline survey the sample included a total of 127 schools and 6304 students, for various reasons (mainly because of school transfers and extended absences due to illness or injuries), there was some attrition by the end of the study. We were able to follow up with 5917 students by the time of the endline survey. A total of 387 students were attrited from the sample. The attrition rate was 6.1%.

According to our analysis, the pattern of attrition did not differ between the treatment and control groups. To understand the nature of those that attrited and assess if the attrition affected the validity of the randomization, we regressed attrition status on the treatment variables. The results comparing the attrition rates between each treatment group and the control group show that the attrition rates are not affected by the treatment status (Table 2). We also regressed the baseline characteristics on treatment variables within non-attrition sample (i.e., students who took part in both the baseline and endline

Table 1
Comparison of the student characteristics between the treatment and control groups within baseline students.

| Variable | Difference between the treatment and control groups within baseline students ^b | |
|---|---|----------------------|
| | CAI treatment (1) | CAL treatment (2) |
| <i>Student characteristics</i> | | |
| [1] Standardized baseline English test score (standard deviation) ^a | 0.02 (0.10) | -0.04 (0.10) |
| [2] Gender (1 = female; 0 = male) | 0.02 (0.02) | 0.00 (0.02) |
| [3] Age (years) | 0.11 (0.10) | 0.03 (0.13) |
| [4] Ethnicity (1 = Han; 0 = minority) | 0.04 (0.06) | 0.04 (0.07) |
| [5] Computer use (1 = ever used computer; 0 = never used computer) | -0.06 (0.07) | -0.01 (0.07) |
| [6] Student self efficacy (1–4 points) | -0.02 (0.03) | 0.05 (0.04) |
| [7] Education level students want to achieve (1 = college or above; 0 = below college) | -0.03 (0.04) | 0.01 (0.05) |
| [8] Selfish (1 = yes; 0 = no) | 0.03 (0.02) | 0.01 (0.02) |
| [9] Altruistic (1 = yes; 0 = no) | -0.05 (0.03) | -0.04 (0.03) |
| [10] Like English teacher (0–100 points) | 2.97 (2.00) | 1.54 (2.27) |
| <i>Family characteristics</i> | | |
| [11] Log(asset) | -0.04 (0.04) | 0.00 (0.03) |
| [12] Only child (1 = yes; 0 = no) | -0.05 (0.04) | 0.02 (0.08) |
| [13] At least one parent doesn't live at home (1 = yes; 0 = no) | -0.04 (0.03) | -0.00 (0.03) |
| [14] At least one parent has junior high school or higher degrees (1 = yes; 0 = no) | -0.04 (0.04) | 0.00 (0.05) |
| [15] Father farms everyday (1 = yes; 0 = no) | 0.00 (0.02) | 0.01 (0.03) |
| [16] Mother farms everyday (1 = yes; 0 = no) | 0.01 (0.03) | -0.01 (0.03) |
| <i>Teacher and school characteristics</i> | | |
| [17] English teacher's gender (1 = female; 0 = male) | -0.09 (0.11) | -0.12 (0.10) |
| [18] English teacher's ethnicity (1 = Han; 0 = minority) | -0.02 (0.09) | 0.02 (0.10) |
| [19] English teacher's age (years) | -0.20 (1.36) | -0.22 (1.32) |
| [20] Whether the English teacher is hired by the education bureau or privately by the school (1 = education bureau; 0 = school) | -0.13 (0.09) | 0.05 (0.10) |
| [21] School area (sq.m) | 2806.37 (2290.90) | 2987.00 (3904.35) |
| [22] Computer-student ratio | 0.26 (0.30) | 0.14 (0.20) |
| [23] Observations | 6304 | 6304 |

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses clustered at school level.

^a The standardized baseline English test score is the score on the standardized English test that is given to all samples students before the ICT program, and it is standardized using the baseline mean and standard deviations for the control group.

^b The within-school difference between the treatment and control groups is calculated by regressions of each of row variables on two treatment dummy variables, controlling for school dummy variables, with robust standard errors clustered at the school level. The sample includes both the non-attrition and the attrition observations.

Source: Authors' survey.

survey and were included in the final analytic sample). None of the regressions showed significant differences between the treatment and control groups (Table 3). In other words, our results show that student characteristics are well balanced between the treatment and control groups both before and after attrition.

2.2. Experiment arm/Intervention

The intervention involved an ICT programme that used computers and a set of learning assisted software to complement the regular English curriculum. Every week, students attended two 40-min programme sessions. The sessions took place during each school's computer class time period. Fifth graders in each of the sample schools all had one computer class that met at least twice per week. Besides using the computers for our CAI and CAL interventions, the students would use computers during computer class to learn basic computer operations, such as how to use a mouse or the keyboard. The programme lasted for the entire 2013–2014 academic year (a total of 28 weeks). A total of 1236 fifth grade students received the CAI treatment and 1068 students received the CAL treatment shortly after the baseline survey at the beginning of the academic year (Fig. 1).

In order to implement either of the computer assisted treatments, we used two pieces of software. First, one set of software for our programme was designed to improve the basic competencies of students in the uniform national English curriculum. The first piece of the software was obtained from a commercial IT company. This software provided both animated reviews and game-based remedial exercises in English for fifth grade students. The content of the software covered all skills and knowledge of the English national curriculum.

Our programme team also developed another piece of the software. The second software package provided a large number of additional exercise questions. We worked together with teachers and experts from the Beijing-based organization of Teaching English to Speakers of Other Languages (TESOL) to choose the questions. The questions were then integrated into an animation-based game interface to make it interesting for the students.

We chose English as our programme subject for several reasons. First, English is one of the main subjects that is used to test students as part of the competitive exam system in China that allows students to compete for positions in high school and college (Bolton & Graddol, 2012; McKay, 2002). In fact, English represents one third of the total points in the high school entrance exam and the college entrance exam. Second, English teaching and English learning is particularly weak in poor rural China (Hu, 2005; Li, 2002; Zhao, 2003). Studies have shown that a low English score is one of the most important reasons that are keeping rural students from attending high school in China (Loyalka, 2014). English teachers are also of notoriously poor quality in rural China (Hu, 2005, 2009). Because of this, we decided to try to use ICT to improve the English teaching/learning in rural schools.

2.2.1. CAI treatment group

In order to facilitate English teachers to incorporate the ICT programme into English teaching practices during CAI classes, we carefully designed and compiled a CAI implementation protocol. The protocol included three parts: a curriculum, a lesson-by-lesson English Teaching Plan (henceforth, the Plan) and a set of instructions on teacher responsibilities. During a one-day intensive training before the programme, we trained the fifth grade English teachers on the implementation protocol. To make sure that help was provided whenever there was a problem about the hardware, the software package or the protocol, we offered "24/7 consultation hotlines" to teachers.

The curriculum (part one of the protocol) was designed to match the progress of English instruction in regular classes. A curriculum instruction manual was given to the supervisor of the programme and this manual included detailed information on a week-by-week basis about the link between each unit in the textbook and the unit in software package. The curriculum also instructed the teachers on how to match specific modules in the software to the pace of learning of students.

The goal of the English Teaching Plan (part two of the protocol) was developed as a way to provide English teachers with a number of interactive activities involving the CAI computer-software package that they were able to use during their classes. The activities not only provided timely feedback to students, they also aimed to increase the interactions between students

Table 2
Comparisons of attrition between the treatment and control groups.

| Dependent variable: attrition (1 = students attrited; 0 = student remained) | |
|---|----------------|
| Variables | Attrition (1) |
| [1] CAI treatment (1 = yes; 0 = no) | −0.01 (0.02) |
| [2] CAL treatment (1 = yes; 0 = no) | −0.01 (0.02) |
| [3] Constant | 0.02*** (0.01) |
| [4] Observations | 6304 |
| [5] R-squared | 0.03 |

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses clustered at school level.

The test aims to show whether attrition rates are different between the treatment and control groups. The test regress attrition status on the treatment variables.

Source: Authors' survey.

Table 3

Comparison of the student characteristics between the treatment and control groups within unattrited students.

| Variable | Difference between the treatment and control groups within unattrited students ^b | |
|---|---|----------------------|
| | CAI treatment (1) | CAL treatment (2) |
| <i>Student characteristics</i> | | |
| [1] Standardized baseline English test score (standard deviation) ^a | 0.02 (0.10) | −0.04 (0.10) |
| [2] Gender (1 = female; 0 = male) | 0.03 (0.02) | 0.01 (0.02) |
| [3] Age (years) | 0.12 (0.10) | 0.01 (0.13) |
| [4] Ethnicity (1 = Han; 0 = minority) | 0.03 (0.06) | 0.04 (0.07) |
| [5] Computer use (1 = ever used computer; 0 = never used computer) | −0.06 (0.07) | −0.00 (0.08) |
| [6] Student self efficacy (1–4 points) | −0.03 (0.03) | 0.05 (0.04) |
| [7] Education level students want to achieve (1 = college or above; 0 = below college) | −0.04 (0.04) | 0.00 (0.05) |
| [8] Selfish (1 = yes; 0 = no) | 0.03 (0.02) | 0.00 (0.02) |
| [9] Altruistic (1 = yes; 0 = no) | −0.05 (0.03) | −0.04 (0.03) |
| [10] Like English teacher (0–100 points) | 2.89 (2.11) | 1.56 (2.36) |
| <i>Family characteristics</i> | | |
| [11] Log(asset) | −0.05 (0.04) | 0.00 (0.03) |
| [12] Only child (1 = yes; 0 = no) | −0.05 (0.05) | 0.02 (0.08) |
| [13] At least one parent doesn't live at home (1 = yes; 0 = no) | −0.03 (0.03) | 0.00 (0.03) |
| [14] At least one parent has junior high school or higher degrees (1 = yes; 0 = no) | −0.04 (0.04) | 0.01 (0.04) |
| [15] Father farms everyday (1 = yes; 0 = no) | 0.01 (0.02) | 0.01 (0.03) |
| [16] Mother farms everyday (1 = yes; 0 = no) | 0.01 (0.03) | −0.01 (0.03) |
| <i>Teacher and school characteristics</i> | | |
| [17] English teacher's gender (1 = female; 0 = male) | −0.09 (0.11) | −0.12 (0.11) |
| [18] English teacher's ethnicity (1 = Han; 0 = minority) | −0.02 (0.09) | 0.02 (0.10) |
| [19] English teacher's age (years) | −0.24 (1.38) | −0.31 (1.32) |
| [20] Whether the English teacher is hired by the education bureau or privately by the school (1 = education bureau; 0 = school) | −0.12 (0.09) | 0.04 (0.10) |
| [21] School area (sq.m) | 2842.24 (2370.76) | 2716.69 (3723.28) |
| [22] Computer-student ratio | 0.28 (0.31) | 0.13 (0.20) |
| [23] Observations | 5917 | 5917 |

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses clustered at school level.

^a The standardized baseline English test score is the score on the standardized English test that is given to all samples students before the ICT programme, and it is standardized using the baseline mean and standard deviations for the control group.

^b The within-school difference between the treatment and control groups is calculated by regressions of each of the row variables on two treatment dummy variables, controlling for school dummy variables, with robust standard errors clustered at the school level. The sample is limited to the unattrited observations.

Source: Authors' survey.

and between students and teachers. In the Plan, each lesson was divided into several teaching sessions. Each session had a learning goal that matched the national curriculum. To meet each learning goals, we listed example of activities that teachers could use to organize the materials or exercises provided by the software package.

In the final (and third) part of the protocol, we laid out the responsibilities that the English teachers needed to fulfill in implementing the CAI classes. These responsibilities included: a) taking attendance of each student at the beginning of each class; b) making sure that the curriculum was followed in a way that each CAI class matched the curriculum of regular English classes; c) organizing activities to enhance interaction and learning; and d) providing timely feedback to students, including providing immediate assistance when students experience difficulty in learning or in computer operations.

To facilitate the implementation of our CAI protocol, we compensated the English teachers who instructed the programme sessions with a stipend of 500 RMB (approximately 80 USD) every semester. We gave the stipend to them by the end of 2013–2014 academic year only if they faithfully implemented the CAI program. To monitor how the English teachers followed the protocol, we recruited enumerators from universities in Haidong Prefecture and sent them to visit the CAI schools during the programme period. The enumerators randomly selected students from each treatment class and surveyed them about how the CAI classes were organized.

2.2.2. CAL treatment group

The most fundamental difference between the CAI treatment group and the CAL treatment group is that the CAL intervention was not created in a way that it tried to become integrated into the English teacher's teaching program. In order to make sure that the supervisors of our intervention did not try to integrate the software into their teaching, non-English teachers were assigned to supervise the CAL classes. The CAL implementation protocol was made up of only a simplified version of the curriculum and a modified version of the teacher responsibilities instruction manual. The CAL supervisors (the teacher in the computer class when the treatment was being given) were still instructed to follow the curriculum as a way to direct the CAL students to the specific unit in the software package (that would match the pace of learning of students) in their

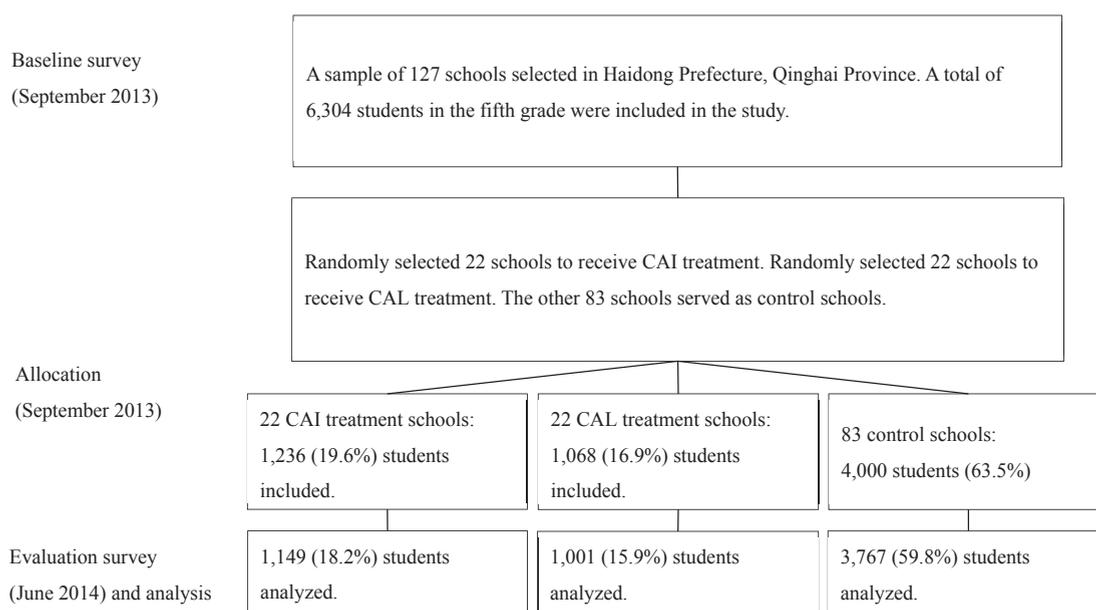


Fig. 1. Experiment profile.

regular English classes. However, in the CAL intervention (unlike the CAI intervention) there was no attempt to integrate the material (and things that students learned) into the regular in-class English curriculum. The CAL protocol also did not include an English Teaching Plan (as did the CAI protocol).

Given this protocol, in a typical CAL session, the students were essentially supervised to use the software package to learn English by themselves. Students watched an animated video that reviewed the materials on which they were receiving instruction during their regular English class sessions. Then students played games to practice the skills introduced in the video lecture. No instruction on English was given during the classes.

In the CAL treatment schools, we also compensated the non-English teachers who instructed the programme sessions with a stipend of 500 RMB (approximately 80 USD) every semester (which made it the same as the compensation to teachers in the CAI treatment arm). The stipend was paid to them by the end of 2013–2014 academic year if they fully implemented the CAL program. We also monitored how the non-English teachers followed the protocol in CAL treatment schools in the same way that we did in the CAI treatment schools.

2.2.3. Control group

A total of 4000 fifth grade students in 83 control schools constituted the control group. During the programme, students in the control group did not receive any intervention. To avoid any form of spillover effects, principals in the control schools (which were all located in separate school districts relative to the treatment schools) were blinded to the experimental nature of our study. To our knowledge, none of principals, teachers or students in the control school were aware of our programme.

2.3. Data collection

The research group conducted two waves of surveys in the 127 sample schools. The first round of survey was a baseline survey conducted with all students in 127 schools in September 2013 at the beginning of the autumn semester. The baseline survey was completed before any implementation of our programme had begun. The second wave of the survey was our evaluation survey, which was conducted at the end of the ICT programme (either the CAI or CAL intervention) in June 2014, a time that coincided with the end of the spring semester of the 2013–2014 academic year.

In each wave of the survey, the enumeration team visited all 127 schools and conducted a three-part survey. In the first part students were given a 30-min standardized English test and we used these scores as our measure of student academic performance. The English test for baseline survey included 25 questions testing listening, 15 testing vocabulary, 8 testing pronunciation, 27 testing grammar and 10 testing dialogues. Before each round of the survey, we tested the English test items with over two hundred 5th grade students to construct baseline and evaluation English exams. All the questions in the English test in the endline were different from the questions in the baseline. We only included questions that did not overlap with the exercises in our software package. We administered and printed the test ourselves to ensure that it was not possible to prepare for the examination. Also, our enumeration team strictly proctored the test, in order to minimize cheating, and enforced the time limit for the exams. We normalized test scores relative to the distribution of the baseline scores of the

control group to make test scores from different rounds of surveying comparable. Specifically, we subtracted the mean of the control group in the baseline and divided by the standard deviation (SD) of the distribution of the control group in the baseline.

In the second part of the survey we collected data on the characteristics of the sample students and their families. From this part of the survey we were able to create a set of demographic and socioeconomic variables. The dataset includes measures of each student's *gender* (1 = female; 0 = male), *age* (years), *ethnicity* (1 = Han; 0 = minority), *computer use* (1 = ever used computer; 0 = never used computer). We also created a number of variables measuring family characteristics, including, assets which is calculated by each account of family durable goods multiplying by their prices, then sum all index and take the logarithm, *only child* (1 = yes; 0 = no), *at least one parent does not live at home* (1 = yes; 0 = no), *at least one parent has junior high school or higher degrees* (1 = yes; 0 = no), *father farms everyday* (1 = yes; 0 = no), and *mother farms everyday* (1 = yes; 0 = no). In the second part of the survey we also asked questions to measure the non-cognitive traits, such as *self-efficacy* (1–4 points), *the education level that students want to achieve* (1 = college or above; 0 = below college), *selfish and altruistic* indicators, and *like English teacher* (0–100 points).

Self-efficacy is a person's perception of their ability to plan and take action to reach a particular goal. "The construct of Perceived Self-Efficacy reflects an optimistic self-belief" (Schwartz, 1992). Perceived self-efficacy is an operative construct, i.e., it is related to subsequent behavior and, therefore, is relevant for clinical practice and behavior change. Jerusalem and Schwartz developed the General Self-Efficacy Scale (GSE) in 1979, which was then widely employed in measuring self-efficacy. GSE has ten items. Each item refers to successful coping and implies an internal-stable attribution of success. In our study, we adopted the Chinese adaption of the GSE developed in Zhang and Schwarzer (1995).

As for *selfish and altruistic* indicators, we followed Jiang (2014) by conducting an experiment to elicit student other-regarding preferences. The experiment asks the students to divide a prize either between herself/himself and other types of people or take the whole prize oneself. There are four different choices that the students can make in an increasing degree of altruism. We generated an altruistic indicator that equals 1 if the student chooses to divide the prize between herself/himself and strangers (the most altruistic choice) and 0 if otherwise. We also generated a selfish indicator if the students choose to take the whole prize herself/himself.

In the third part of the survey, we gathered information about English teachers and schools. For example, we had questions about *English teacher's gender* (1 = female; 0 = male), *English teacher's ethnicity* (1 = Han; 0 = minority), *English teacher's age* (years), and *whether the English teacher is hired by the education bureau or privately by the school* (1 = education bureau; 0 = school). If a teacher is hired by the education bureau, he or she is a registered public teacher. The rural schools only have the right to hire contract teachers that cannot be registered in the education bureau. A registered teacher hired by education bureau gets higher basic salary and subsidies than the privately hired contract teachers. We also collected information on the basic characteristics of the schools, such as *school area* (sq. m) and *computer-student ratio*.

2.4. Statistical methods

To evaluate how academic performance changed among students who participated in the CAI or CAL programme relative to those who did not, we used both unadjusted and adjusted ordinary least squares (OLS) regression analysis. The regression models are presented in order of increasing comprehensiveness.

To test the impact of our programme, we estimated Average Treatment Effect (ATE) by regressing the endline outcome variable (i.e. *standardized endline English test score*) on dummy variables of the treatment status. The model we estimated is:

$$y_{is} = \alpha + \beta_{CAL} \cdot treatment_{1s} + \beta_{CAI} \cdot treatment_{2s} + \theta \cdot y_{0is} + \varepsilon_{is} \quad (1)$$

where y_{is} is the standardized endline English test score for child i in school s , $treatment_{1s}$ is the dummy variable for the CAL programme and $treatment_{2s}$ is the dummy variable for the CAI programme. Because the treatment was randomly assigned, β_{CAL} and β_{CAI} in Equation (1) provide unbiased estimates of the average treatment effect of the CAL programme and the CAI programme respectively. We also included y_{0is} , which is the English test score of the same child at the baseline. Therefore, the coefficients, β_{CAL} and β_{CAI} , measure the changes in English test scores between the periods before and after programme implementation that can be attributed to the treatment.

To enhance the efficiency of estimation, we estimated an adjusted model by controlling for a set of variables collected at the baseline survey. The model we estimated is:

$$y_{is} = \alpha + \beta_{CAL} \cdot treatment_{1s} + \beta_{CAI} \cdot treatment_{2s} + \theta \cdot y_{0is} + X_{is} \gamma + \varepsilon_{is} \quad (2)$$

where X_{is} includes a vector of covariates of student, family and school characteristics (such as gender, age, ethnicity, computer use, education level students want to achieve, assets, characteristics of English teacher, school area, computer-student ratio, etc.)

In order to test for heterogeneous programme effects (that is, whether certain types of ICT programmes work best for certain types of students/teachers), we also included interaction terms between the treatment dummy variables and some key covariates in the regression model specified in Equation (2). Specifically, we will test whether the change in English test scores before and after the programme differ for students who were performing better or worse in English at the time of the

baseline survey and we will do this for each of two ICT programmes, CAI and CAL. We do this by including in the regression an interaction term between the treatment dummy variables and the variable measuring the baseline English test score.

The model we estimated is:

$$y_{is} = \alpha + \beta_{CAL1} \cdot treatment_{1s} + \beta_{CAI1} \cdot treatment_{2s} + \beta_{CAL2} \cdot treatment_{1s} \cdot y_{0is} + \beta_{CAI2} \cdot treatment_{2s} \cdot y_{0is} + \theta \cdot y_{0is} + X_{is} \gamma + \varepsilon_{is} \quad (3)$$

where the coefficients on the interaction terms, β_{CAL2} and β_{CAI2} , indicate the heterogeneous treatment effects by student baseline English test score. In all regressions, we accounted for the clustered design by constructing Huber-White standard errors clustered on the school level (relaxing the assumption that disturbance terms are independent and identically distributed within schools).

We also use a version of Equation (3) to test whether the changes in English test scores of students before and after each of the two ICT programmes differ depending on types of teachers. We do this by including an interaction term between the treatment dummy variables and the variable measuring the workload of the ICT programme supervisors in the regression. The model to test this is:

$$y_{is} = \alpha + \beta_{CAL1} \cdot treatment_{1s} + \beta_{CAI1} \cdot treatment_{2s} + \beta_{CAL2} \cdot treatment_{1s} \cdot workload_{CAL} + \beta_{CAI2} \cdot treatment_{2s} \cdot workload_{CAL} + \theta \cdot y_{0is} + X_{is} \gamma + \varepsilon_{is} \quad (4)$$

This regression is estimated in a way similar to the way in which we estimated Equation (3).

3. Results

In this section we seek to understand how computers and software can most effectively be used in classrooms to raise the educational performance of children. To do this, we first combine the data from the CAI and CAL treatments into one mega-treatment group. This allows us to test if ICT affects student outcomes without taking into consideration the way ICT is used in the classroom. Second, as shown in Equations (1) and (2), we test the effect of the different treatments (CAI and CAL) on student test scores in order to examine the value of integrating ICT into teaching methods and curriculums. Finally, we examine whether the two approaches to using ICT in the classroom have differential impacts on different types of students or whether the success of ICT integration relies on the nature of a teacher.

3.1. The average impact of any ICT programme

According to our data, when we combine the two treatment groups and test the average impact of any ICT programme on student English test scores, we find there is no significant impact on the English performance of students (Table 4). Both the unadjusted model and adjusted model show that although the point estimate of the programme impact is estimated to be 0.05 standard deviations, it is not statistically significant (column 1, rows 1 and 2). Therefore, when combining the two types of ICT-based programmes (both CAI and CAL), our results suggest that there is no detectable impact on student educational performance.

The results change, however, when we separate the two treatment groups (CAI and CAL). Based on the results produced by estimating Equations (1) and (2), we find that the CAI programme is effective in raising test scores while the CAL programme is not (Table 5). Specifically, the CAI treatment (which integrates ICT with teaching) is estimated to raise test scores by 0.07

Table 4

Ordinary Least Squares estimators of the impact of ICT programme on standardized English test scores.

| Dependent Variable: Standardized endline English test score (standard deviation) | | |
|--|----------------------|--------------------|
| | (1) Unadjusted Model | (2) Adjusted Model |
| [1] Any treatment (1 = yes; 0 = no) | 0.05 (0.04) | 0.05 (0.04) |
| [2] Standardized baseline English test score (standard deviation) ^a | 0.71*** (0.03) | 0.68*** (0.03) |
| [3] Controls | NO | YES |
| [4] Observations | 5917 | 5917 |
| [5] R-squared | 0.72 | 0.73 |

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses clustered at school level.

The test aims to show the impact of the treatment on student English test scores, the test regresses standardized endline English test scores on treatment variable for unadjusted model and regresses standardized endline English test scores on both treatment variables and a set of control variables for adjusted model.

^a The standardized baseline English test score is the score on the standardized English test that is given to all samples students before the ICT programme, and it is standardized using the baseline mean and standard deviations for the control group.

Source: Authors' survey.

Table 5
Ordinary Least Squares estimators of the impact of ICT programme on standardized English test scores.

| Dependent Variable: Standardized endline English test score (standard deviation) | | |
|--|----------------------|--------------------|
| | (1) Unadjusted Model | (2) Adjusted Model |
| [1] CAI treatment (1 = yes; 0 = no) | 0.07* (0.04) | 0.08** (0.04) |
| [2] CAL treatment (1 = yes; 0 = no) | 0.03 (0.06) | 0.03 (0.05) |
| [3] Standardized baseline English test score (standard deviation) ^a | 0.71*** (0.03) | 0.68*** (0.03) |
| [4] Controls | NO | YES |
| [5] Observations | 5917 | 5917 |
| [6] R-squared | 0.72 | 0.73 |

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses clustered at school level.

The test aims to show the impact of the treatment on student English test scores, the test regresses standardized endline English test scores on the treatment variables and a set of control variables.

^a The standardized baseline English test score is the score on the standardized English test that is given to all samples students before the ICT programme, and it is standardized using the baseline mean and standard deviations for the control group.

Source: Authors' survey.

standard deviations. This impact estimate is significant at the 10% level using the unadjusted model (column 1, row 1). After controlling for our set of specified baseline covariates, the impact slightly increases to 0.08 standard deviations. The estimate from the adjusted model is significant at the 5% level (column 2, row 1).

In contrast, the results from estimating Equations (1) and (2) find that the CAL programme (which was not integrated into a classroom's teaching programme) was not effective. The coefficient on the CAL treatment, while positive (0.03 standard deviations), is not statistically significant (columns 1 and 2, row 2). In addition, the difference between the coefficients on the CAI and the CAL treatment variables is significant at the 1% level. Hence, our findings show that the programme impacts differ between CAI and CAL treatments. This result is consistent with an interpretation that an ICT programme is more effective when it is integrated into a classroom's teaching programme.

3.2. Heterogeneous treatment effects

In order to examine whether either type of ICT programmes (CAI or CAL) is most effective at helping certain types of students, we used Equation (3) to test the heterogeneous effects of the two treatments by incorporating the initial levels of educational performance of the students into our analysis. When we interact the CAI treatment variable with the variable measuring student English academic performance at the baseline, the coefficient on the interaction term is small in magnitude and insignificant (Table 6, column 1, row 2). This finding suggests that better performing and poorer performing students benefit similarly from the CAI treatment. However, when we interact the CAL treatment variable with baseline English scores, we find that the coefficient on the interaction term is 0.08 standard deviations and is significant at the 10% level (column 1, row 4). In contrast to the results for the CAI programme, these findings suggest that the CAL programme works better on the better performing students than on poorer performing ones.

So what do such heterogeneous results imply? The results are consistent with the literature suggesting that the effectiveness of an ICT programme depends on a student's initial level of knowledge on a particular subject (Gass & Mackey, 2006;

Table 6

Ordinary Least Squares estimators of the heterogeneous impact of ICT programme on standardized English test scores of students baseline score.

| Dependent Variable: Standardized endline English test score (standard deviation) | |
|--|----------------|
| | (1) |
| [1] CAI treatment (1 = yes; 0 = no) | 0.07* (0.04) |
| [2] CAI treatment* standardized baseline English test score | 0.03 (0.04) |
| [3] CAL treatment (1 = yes; 0 = no) | 0.03 (0.05) |
| [4] CAL treatment* standardized baseline English test score | 0.08* (0.04) |
| [5] Standardized baseline English test score (standard deviation) ^a | 0.66*** (0.03) |
| [6] Controls | YES |
| [7] Observations | 5917 |
| [8] R-squared | 0.73 |

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses clustered at school level.

The test aims to show the heterogeneous impact of the ICT programme on standardized English test scores of students baseline score.

^a The Standardized baseline English test score is the score on the standardized English test that is given to all samples students before the ICT programme, and it is standardized using the baseline mean and standard deviations for the control group.

Source: Authors' survey.

Hatch, 1978; Long, 1996; Sauro, 2009). In the case of students who performed better at baseline, their level of knowledge of English may have been high enough for them to benefit from the CAL programme. However, students who performed worse on the baseline assessment may not have been able to benefit from the ICT programme without more focused and student-specific instruction from an English teacher. Such an interpretation suggests that when teachers provide guidance and corrective feedback (as was the case in our CAI program) students who initially perform poorly are able to benefit from ICT programming. Hence, unless one can be sure that a CAL programme is appropriate for a classroom's learning level, CAI-type programmes would be more effective at instructing every student in a class—especially when introducing a new subject.

Using Equation (4), we were also able to test whether the workload of the supervisor of our ICT programmes at baseline had an effect on the success of either of the two programmes. Our results show that the effectiveness of a CAI programme is not heavily impacted by the workload of a supervisor (Table 7). The coefficient on the interaction term between the variables indicating CAI treatment and supervisor's workload is statistically insignificant (column 1, row 3). However, in the case of the CAL treatment group, the workload of the supervisor appears to matter. If the supervisor's teaching workload is lower than the mean, the CAL programme improved the English performance of students by 0.15 standard deviations; the coefficient was significant at the 10% level (column 1, row 4). In contrast, if the supervisor had a high workload, the CAL programme had no impact (column 1, row 6). Hence, this finding suggests that a teacher's workload and teaching activities complement a CAI programme when it is incorporated in regular in-school English instruction. However, the same is not true for the CAL treatment. There does not appear to be any complementarities between CAL programmes and the work activities of teachers within our findings.

So what is happening? Are certain ICT programmes more or less effective for particular types of teachers? Our results suggest that the potential effects of teacher workload are expected to manifest differently for teachers implementing either CAI or CAL program. Because the supervisors of CAI programmes are also English teachers, implementing CAI programmes clearly complements and increases the efficiency of their teaching, subsequently decreasing their overall workload. Hence, it is possible that CAI supervisors are more likely to fully implement an ICT programme even when they have a heavy workload. This finding is also consistent with the suggestion of Lam (2000) and Nomass (2013) that if an ICT programme complements and increases the efficiency of teaching, the programme may decrease the overall effort or workload of teachers. In contrast, since implementing a CAL programme leads to an absolute increase in the overall workload of a CAL supervisor, when a CAL supervisor has a heavy regular school workload, he/she may be less willing to fully implement the programme.

4. Conclusion

In this paper we investigated different methods of designing and using ICT in schools to effectively raise students' level of educational performance. After showing that, in general, ICT programmes work, we then designed a study to compare the relative effectiveness of ICT programmes that are either integrated into the teaching programmes of the classrooms in schools (CAI treatments) or not (CAL treatments). We also sought to show that not all ICT programmes are appropriate for all classroom situations. Specifically, we explored whether one type of ICT programme is most effective for certain types of students and teachers.

To meet these objectives, we conducted a cluster-randomized controlled trial of an ICT programme that aimed to improve student learning in English in 127 rural public schools in Haidong prefecture in Qinghai province. This study involved 6304

Table 7

Ordinary Least Squares estimators of the heterogeneous impact of ICT programme on standardized English test scores of the supervisor's teaching workload.

| Dependent Variable: Standardized endline English test score (standard deviation) | |
|--|----------------|
| | (1) |
| [1] CAI treatment (1 = yes; 0 = no) | 0.10** (0.05) |
| [2] CAI supervisor's workload (1 = yes; 0 = no) | -0.00 (0.05) |
| [3] CAI treatment* CAI supervisor's workload | -0.06 (0.08) |
| [4] CAL treatment (1 = yes; 0 = no) | 0.15* (0.08) |
| [5] CAL supervisor's workload (1 = yes; 0 = no) | 0.02 (0.04) |
| [6] CAL treatment* CAL supervisor's workload | -0.18* (0.11) |
| [7] Standardized baseline English test score (standard deviation) ^a | 0.67*** (0.03) |
| [8] Controls | YES |
| [9] Observations | 5917 |
| [10] R-squared | 0.73 |

* significant at 10%; ** significant at 5%; *** significant at 1%. Robust standard errors in parentheses clustered at school level.

The test aims to show the heterogeneous impact of ICT programme on standardized English test scores of the supervisors' workload.

^a The standardized baseline English test score is the score on the standardized English test that is given to all samples students before the ICT programme, and it is standardized using the baseline mean and standard deviations for the control group.

Source: Authors' survey.

fifth grade students. To our knowledge, this is the largest experiment ever done to evaluate the effectiveness of a programme that integrates an ICT programme into the teaching curriculum of a classroom environment (a CAI programme). In addition, this study was designed to compare a CAI intervention with a programme that did not integrate ICT into a teacher's teaching curriculum (a CAL programme).

The results show that there are significant differences in programme impacts depending on whether an ICT programme is integrated into teacher's ordinary curriculum (CAI) or not (CAL). We believe that students in the sample schools benefited from the ICT programme most when teachers integrated it into their teaching programme and provided instructions to guide students in using its materials. When the ICT programme was not integrated into the teaching programme, however, there was no improvement in student English scores.

The final part of the study sought to explain how differences in programme impacts between different types of ICT programmes might emerge. We believe the finding that CAI programmes had larger impact than CAL programmes can be partly explained by differences in students' initial achievement levels within our sample. The effectiveness of CAL programming was very different for who were either high or low achieving at baseline – students with higher levels of initial learning benefitted, while those with lower levels did not. However, when teachers implemented a CAI programme by integrating the ICT programme into their teaching curriculum and providing additional instruction, both types of students improved.

Our analysis also uncovered potential complementarities between types of teachers and ICT programmes. The ICT programme was shown to be effective regardless of teacher's regular workload if it was an English teacher who integrated the programme into their teaching programme (CAI programs). However, if a non-English teacher implemented the ICT programme (CAL programs), the programme was not effective if the teacher's workload was heavy.

These results have several implications for how an ICT programme might be designed and implemented in rural China or other developing countries. First, when designing the ICT programme, attention needs to be paid to the learning levels of the targeted students. If the subject being taught by the ICT programme is a new subject that students have just started to learn or one in which students are lagging behind, an ICT programme may be more effective if teachers integrate it into their teaching program. If it is a subject in which students have established a good foundation of knowledge, they may be able to follow the instructions of the learning assisted software and engage in self-study on their own. In the latter case, it may be less necessary to integrate the ICT programme into teaching programs.

Second, in the case of CAL programmes where the ICT programme is designed to be delivered without being integrated into a teaching programme (and the programme supervisor themselves will not teach the subject), assigning teachers with high workloads (or overburdened teachers) to the position of supervisor may undermine the effectiveness of the programme. When there are not strong complementarities between the ICT programme and the supervisor's regular teaching, supervisors with heavy workloads may choose to not make the necessary effort to fully implement the programme sessions. Therefore, CAL programmes would be much more effective if the supervisor is not overloaded with other teaching duties and has the necessary capacity to faithfully implement the ICT programme.

Hence, according to our study, not all ICT programmes are equal. While ICT programmes work in general, different programmes—at least in our study areas—work differently in different environments (with different types of students; and with different types of teachers). Of course, this heterogeneity may help explain the pattern of outcomes of previous ICT studies in which certain programmes were found to affect educational performance and others were not. In this study when we run a large-scale experiment with different types of ICT programmes, we find that there can be different impacts depending on the nature of the study and the nature of the participants. This finding suggests that future work in this area may be necessary in order to identify additional sets of factors that affect the effectiveness of ICT programmes.

Acknowledgments

The authors acknowledge the financial assistance of the National Natural Science Foundation of China (grant numbers 71473240 and 71333012) and the International Initiative for Impact Evaluation (3ie).

References

- Almekhlafi, A. G. (2006). The effect of computer assisted language learning (CALL) on United Arab Emirates English as a foreign language (EFL) school students' achievement and attitude. *Journal of Interactive Learning Research*, 17(2), 121–142.
- Angrist, J., & Lavy, V. (2002). New evidence on classroom computers and pupil learning*. *The Economic Journal*, 112(482), 735–765.
- Bakar, A., Inal, Y., & Cagiltay, K. (2006). Use of commercial games for educational purposes: will today's teacher candidates use them in the future?. In *World conference on educational multimedia, hypermedia and telecommunications* (Vol. 2006(1), pp. 1757–1762).
- Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: evidence from two randomized experiments in India. *Quarterly Journal of Economics*, 122(3), 1235–1264.
- Barrow, L., Markman, L., & Rouse, C. E. (2008). *Technology's edge: The educational benefits of computer-aided instruction* (No. w14240). National Bureau of Economic Research.
- Bingimlas, K. A. (2009). Barriers to the successful integration of ICT in teaching and learning environments: a review of the literature. *Eurasia Journal of Mathematics, Science & Technology Education*, 5(3), 235–245.
- Bolton, K., & Graddol, D. (2012). English in China today. *English Today*, 28(03), 3–9.
- Burston, J. (1996). CALL at the crossroads: myths, realities, promises, and challenges. *Australian Review of Applied Linguistics*, 19(2), 27–36.

- Butcher, N. (2011). *ICT, education, development, and the knowledge society. Thematic paper prepared for the global e-schools and communities initiative (GeSCI) African leadership in ICT program* (Vol. 81, p. 29). Retrieved from <http://www.gesci.org/assets/files/ICT%20Education%20Development%20and%20the%20Knowledge%20Society>.
- Chien, Y. C. (2011). *Effects of computer-assisted language learning (CALL) instruction on the acquisition of passive grammatical forms by post-secondary English as a second language (ESL) students*. Doctoral dissertation. Florida: University of Central Florida Orlando.
- Dina, A. T., & Ciornei, S. I. (2013). The advantages and disadvantages of computer assisted language learning and teaching for foreign languages. *Procedia-Social and Behavioral Sciences*, 76, 248–252.
- Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: a toolkit. *Handbook of development economics*, 4, 3895–3962.
- Dynarski, M., Agodini, R., Heavyside, S., Novak, T., Carey, N., Campuzano, L., et al. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort*. Report to Congress. National Center for Education Evaluation and Regional Assistance.
- Ebner, M., & Holzinger, A. (2007). Successful implementation of user-centered game based learning in higher education: an example from civil engineering. *Computers & Education*, 49(3), 873–890.
- Gass, S. M., & Mackey, A. (2006). Input, interaction and output: an overview. *AILA review*, 19(1), 3–17.
- Granger, C. A., Morbey, M. L., Lotherington, H., Owston, R. D., & Wideman, H. H. (2002). Factors contributing to teachers' successful implementation of IT. *Journal of Computer Assisted Learning*, 18(4), 480–488.
- Hatch, E. M. (1978). *Second language acquisition: A book of readings*. Newbury House Pub.
- He, F., Linden, L., & MacLeod, M. (2008). *How to teach English in India: Testing the relative productivity of instruction methods within the Pratham English language education program*. New York, United States: Columbia University. Mimeographed document.
- Hismanoglu, M. (2012). Prospective EFL teachers' perceptions of ICT integration: a study of distance higher education in Turkey. *Educational Technology & Society*, 15(1), 185–196.
- Hlas, A. C., & Vuksanovich, M. (2007). Computer assisted language learning in Spanish elementary school foreign language classrooms: the role of CD-ROMS. *Hispania*, 769–783.
- Hu, G. (2005). English language education in China: policies, progress, and problems. *Language Policy*, 4(1), 5–24.
- Hu, X. (2009). The quality of English teacher of the primary school in rural China—evidence from Heilongjiang Province (in Chinese). *China Adult Education*, 18, 89–90.
- Huang, W., Di, M., Yaojiang, S., Linxiu, Z., Matthew, B., & Scott, R. (2014). *Computer technology in Education: Evidence from a pooled study of computer assisted learning programmes among rural students in China*. REAP working paper. Retrieved from: <http://reap.fsi.stanford.edu/publication/computer-technology-education-evidence-pooled-study-computer-assisted-learning-programs>.
- Hughes, J. (2005). The role of teacher knowledge and learning experiences in forming technology-integrated pedagogy. *Journal of Technology and Teacher Education*, 13(2), 277–302.
- Jiang, T. (2014). *Other-regarding preferences and other-regarding cheating—experimental evidence from China, Italy, Japan and the Netherlands*. Italy, Japan and the Netherlands (December 1, 2014).
- Jones, J. F. (2001). CALL and the responsibilities of teachers and administrators. *ELT Journal*, 55(4), 360–367.
- Krashen, S. (1982). *Principles and practice in second language acquisition* (pp. 1982–1982). Pergamon: Oxford.
- Lai, F., Luo, R., Zhang, L., Huang, X., & Rozelle, S. (2011). *Does computer-assisted learning improve learning outcomes? Evidence from a randomized experiment in migrant schools in Beijing*. REAP working paper.
- Lai, F., Zhang, L., Hu, X., Qu, Q., Shi, Y., Qiao, Y., et al. (2013). Computer assisted learning as extracurricular tutor? Evidence from a randomised experiment in rural boarding schools in Shaanxi. *Journal of Development Effectiveness*, 5(2), 208–231.
- Lai, F., Zhang, L., Qu, Q., Hu, X., Shi, Y., Boswell, M., et al. (2012). *Does computer-assisted learning improve learning outcomes? Evidence from a randomized experiment in public schools in rural minority areas in Qinghai, China*. REAP working paper.
- Lam, Y. (2000). Technophilia vs. technophobia: a preliminary look at why second-language teachers do or do not use technology in their classrooms. *Canadian Modern Language Review*, 56(3), 390–420.
- Li, S. (2002). Status and development of English teaching in minority areas (in Chinese). *Foreign Language Teaching & Research in Basic Education*, 11, 34–39.
- Linden, L. L. (2008). *Complement or substitute?: The effect of technology on student achievement in India*. InfoDev.
- Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie, & T. K. Bathia (Eds.), *Handbook of research on second language* (pp. 413–468). San Diego, CA: Academic Press.
- Loyalka, P., Shi, Z., Chu, J., Johnson, N., Wei, J., & Rozelle, S. (2014). *Is the high school admissions process fair? Explaining inequalities in elite high school enrollments in developing countries*. REAP working paper.
- McKay, S. L. (2002). *Teaching English as an international language: Rethinking goals and perspectives*. NY: OUP.
- Ministry of Education. (2001). *The notice of "compulsory education curriculum pilot program" from the ministry of education*. http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/moe_711/201006/xxgk_88602.html.
- Ministry of Education. (2012). *Notice about "accelerating the process of information technology integration into education"*. Ministry of Education.
- Ministry of Education. (2012). *Ten-year development plan for integrating information technology into education*. Retrieved March 13, 2012, from http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/s5892/201203/xxgk_133322.html.
- Ministry of Education. (2012). *12th five-year plan for integrating information technology into education*.
- Ministry of Education. (2014). *Notice of "working priorities for integrating information technology into education"*. Ministry of Education. Retrieved March 12, 2014, from <http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/s7062/201403/165870.html>.
- Mo, D., Zhang, L., Luo, R., Qu, Q., Huang, W., Wang, J., et al. (2014). Integrating computer-assisted learning into a regular curriculum: evidence from a randomised experiment in rural schools in Shaanxi. *Journal of Development Effectiveness*, 6(3), 300–323.
- Mo, D., Zhang, L., Wang, J., Huang, W., Shi, Y., Boswell, M., et al. (2014). *The persistence of gains in learning from computer assisted learning (CAL): Evidence from a randomized experiment in rural schools in Shaanxi province in China*. REAP working paper.
- Mumtaz, S. (2000). Factors affecting teachers' use of information and communications technology: a review of the literature. *Journal of Information Technology for Teacher Education*, 9(3), 319–342.
- National Bureau of Statistics of China (NBS). (2014). *China national statistical yearbook, 2014*. Beijing, China: China State Statistical Press.
- Nomass, B. B. (2013). The impact of using technology in teaching English as a second language. *English Language and Literature Studies*, 3(1), p111.
- Olibie, E. I. (2010). Using computer-assisted language learning to improve students' English language achievement in universal basic education. *International Journal of Educational Research and Technology*, 1(1), 66–71.
- Pal, J., Pawar, U. S., Brewer, E. A., & Toyama, K. (2006, May). The case for multi-user design for computer aided learning in developing regions. In *Proceedings of the 15th international conference on World Wide Web* (pp. 781–789). ACM.
- Rahimi, M., & Yadollahi, S. (2011). ICT use in EFL classes: a focus on EFL teachers' characteristics. *World Journal of English Language*, 1(2), p17.
- Rouse, C. E., & Krueger, A. B. (2004). Putting computerized instruction to the test: a randomized evaluation of a "scientifically based" reading program. *Economics of Education Review*, 23(4), 323–338.
- Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning & Technology*, 13(1), 96–120.
- Schwartz, R. (1992). *Self-efficacy in the adoption and maintenance of health behaviours: Theoretical approaches and a new model. Self-Efficacy: Thought control of action* (pp. 217–243). Washington, DC: Hemisphere.
- Stockwell, G. (2007). A review of technology choice for teaching language skills and areas in the CALL literature. *ReCALL*, 19(02), 105–120.
- Sweet, R., & Meates, A. (2004). ICT and low achievers: what does PISA tell us? *Promoting Equity Through ICT in Education: Projects, Problems, Prospects*, 13.

- Tondeur, J., Valcke, M., & Van Braak, J. (2008). A multidimensional approach to determinants of computer use in primary education: teacher and school characteristics. *Journal of Computer Assisted Learning*, 24(6), 494–506.
- Yang, S. C., & Huang, Y. F. (2008). A study of high school English teachers' behavior, concerns and beliefs in integrating information technology into English instruction. *Computers in Human Behavior*, 24(3), 1085–1103.
- Zhang, J. X., & Schwarzer, R. (1995). Measuring optimistic self-beliefs: a Chinese adaptation of the general self-efficacy scale. *Psychologia*, 38(3), 174–181.
- Zhang, J., Zhang, X., Zhao, J., & Wu, D. (2014). *The international development report on ICT in education (2013-2014)*. Beijing Normal University Publishing Group (in Chinese).
- Zhao, H. (2003). Exploring resources of local courses an effective way of developing teaching body in ethnic minority regions. *Journal of Research on Education for Ethnic Minorities*, 2003(4), 60–64.
- Zheng, H. (2009). English teaching and learning in rural China primary school (in Chinese). *Journal of Teaching and Management*, 12, 82–83.